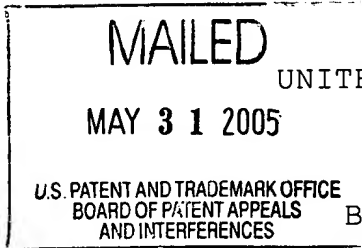


The opinion in support of the decision being entered today was not written for publication in a law journal and is not binding precedent of the Board.

Paper No. 21



UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

Ex parte MICHAEL E. PALMER, GORDON SUN
and HONGYUAN ZHA

Appeal No. 2005-0481
Application No. 09/333,121

ON BRIEF

Before KRASS, BARRY and SAADAT, Administrative Patent Judges.
KRASS, Administrative Patent Judge.

DECISION ON APPEAL

This is a decision on appeal from the final rejection of claims 1-34.

The invention pertains to the examination of electronic documents. In particular, a data processor determines if one electronic document is similar to another among a large collection of such documents that may be maintained in an indexing, or search and retrieval system.

Representative independent claim 1 is reproduced as follows:

1. A method of categorizing a plurality of new electronic documents into a set of categories, comprising the steps of:

establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;

determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories by determining similarity between said each document and the training documents that belong to the training set of said category; and

wherein the step of determining similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity.

The examiner relies on the following references:

Pirolli et al. (Pirolli)	5,835,905	Nov. 10, 1998
Prasad	5,960,422	Sep. 28, 1999
		(filed Nov. 26, 1997)
Bengio et al. (Bengio)	6,128,606	Oct. 3, 2000
		(filed Mar. 11, 1997)
Hoffert et al. (Hoffert)	6,282,549	Aug. 28, 2001
		(filed Mar. 29, 1999)
Chakrabarti et al.	6,389,436	May 14, 2002
(Chakrabarti)		(filed Dec. 15, 1997)

Claims 1-34 stand rejected under 35 U.S.C. § 103. As evidence of obviousness, the examiner offers Pirolli and Prasad with regard to claims 1-14, 17-19, and 26-34, alternatively adding Hoffert with regard to claims 15 and 16, Bengio with regard to claim 20, and Chakrabarti with regard to claims 21-25.

Appeal No. 2005-0481
Application No. 09/333,121

Reference is made to the brief and answer for the respective positions of appellants and the examiner.

OPINION

With regard to the independent claims, it is the examiner's view that Pirolli discloses the claimed subject matter but for establishing a plurality of training sets wherein each training set is associated with a category and includes training documents that have been classified as belonging to the associated category, and for determining similarity between each document and the training documents.

The examiner turns to Prasad, at column 4, lines 17-21, for a teaching of using a set of training documents, and at column 2, lines 50-67, for a suggestion that this method would enable searches more likely to satisfy a user query, and concludes that it would have been obvious to have established a plurality of training sets in Pirolli, wherein each training set is associated with a category and includes training documents that have been classified as belonging to the associated category.

Moreover, the examiner contends that it would have been obvious to determine similarity between documents and documents in the training set because artisans

Appeal No. 2005-0481
Application No. 09/333,121

[W]ould have recognized that using sets of training documents to automatically define categories would have provided the benefit of basing criteria on documents actually taken from the category which, Prasad explains . . . would have provided more accurate search results. Therefore, it would have been obvious . . . to extend Pirolli with Prasad's teaching of using training documents [answer, page 4).

Appellants argue that Pirolli compares documents to rules and not to documents and that Prasad applies rule induction to the training set to first generate rules, and then the rules, not the training documents, are used to determine to what source to direct queries. We are unpersuaded by this argument. It is true that Pirolli generates a set of rules and then compares documents to those rules for a measure of similarity. However, the instant claims do not preclude this arrangement. Pirolli does categorize documents according to classification characteristics and determines similarity between documents in order to categorize a document. Prasad employs a set of training documents and certain rules are applied to the training set. These rules are used to determine similarity between a training set of documents and documents to be classified. While Prasad and/or Pirolli may compare documents to rules, the end result is a comparison of documents to a training set of documents because both types of documents are compared to, and subject to, the same set of rules. The instant claims do not preclude this possibility.

Appellants argue that the combination of Pirolli and Prasad is improper because they both use a comparison of rules to documents rather than documents to documents for categorizing documents or directing queries (brief-page 18) and therefore the combination cannot suggest comparing a plurality of documents to sets of documents to categorize the documents. We are unpersuaded for the reason supra, i.e., the use of rules as an intermediary in the comparison of documents is not precluded by the instant claims.

Appellants argue that the activation of Pirolli is not a sub-step of categorizing. That is, instant claims 1 and 34 require the "determining similarity . . ." to be a substep of the step of "determining how strongly . . ." because the "determining how strongly . . ." step is made possible by "determining similarity . . ." By contrast, argue appellants, Pirolli's categorization steps are preparation for the spreading activation and are not performed by the spreading activation. Yet, the examiner associates Pirolli's categorization of the training set with "determining how strongly each document . . . corresponds to each of said plurality of categories" and the examiner associates Pirolli's spreading activation with "determining

Appeal No. 2005-0481
Application No. 09/333,121

similarity . . ." Thus, argue appellants, the examiner's explanation is inconsistent with the specific claim language.

The examiner's response is to point to column 8, lines 8-47, of Pirolli for an alleged disclosure of determining how strongly each document corresponds to each of the categories by determining similarity between each document and a set of criteria established for the category (answer-page 16).

We will not sustain the rejection of claims 1 and 34 under 35 U.S.C. § 103 because, in our view, the examiner has not established a prima facie case of obviousness with regard to the claimed subject matter.

The portion of Pirolli (column 8, lines 8-47) relied on by the examiner for a showing of the claimed, "determining how strongly each document . . . corresponds to each of a plurality of categories . . .," is directed to performing categorizations by using a vector matrix and weighted linear equations that define the rules for predicting degree of category membership for each page of a web locality. Yet, the instant claims require that the determination of how strongly each document corresponds to each of a plurality of categories is performed "by determining similarity between said each document and the training

documents . . ." This is much different than anything taught by Pirolli. There is no indication in Pirolli that the weighted linear equations used for predicting a "degree of category membership" (column 8, line 43) is a determination of similarity between said each document and the training documents, as claimed.

The instant claims also require that the step of determining similarity is performed by using a matrix representing document similarity derived by combining two or more measures of document similarity. It is true that Pirolli mentions that vectors of features constructed from "text similarities" (column 8, line 11) are collected into a matrix, but it appears that Pirolli determines how strongly documents correspond by the use of weighted linear equations, with the result being entered as part of the vector matrix, rather than using the matrix to determine how strongly documents correspond, as required by the instant claims.

Accordingly, since an important part of the claimed subject matter, i.e., determining how strongly each document corresponds to each of a plurality of categories by determining, using a matrix, similarity between each document and the training

Appeal No. 2005-0481
Application No. 09/333,121

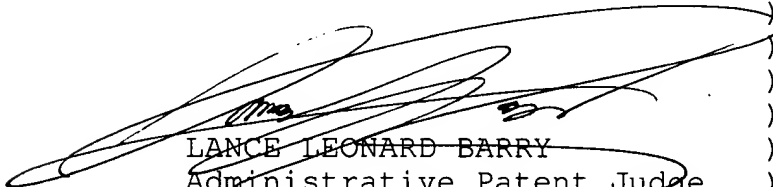
documents, is not taught or suggested by the applied references,
we will not sustain the rejection of claims 1-34 under 35 U.S.C.
§ 103.

The examiner's decision is reversed.

REVERSED



ERROL A. KRASS)
Administrative Patent Judge)



LANCE LEONARD BARRY)
Administrative Patent Judge)

BOARD OF PATENT
APPEALS AND
INTERFERENCES



MAHSHID D. SAADAT)
Administrative Patent Judge)

EAK:clm

Appeal No. 2005-0481
Application No. 09/333,121

Hickman, Palermo, Truong & Becker, LLP
2055 Gateway Place
Suite 550
San Jose, CA 95110